Low-resolution Character Recognition by Video-based Super-resolution

Ataru Ohkura[1], Daisuke Deguchi[1], Tomokazu Takahashi[2], Ichiro Ide[1] and Hiroshi Murase[1]

[1] Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan

okura@murase.m.is.nagoya-u.ac.jp, {ddeguchi, ide, murase}@is.nagoya-u.ac.jp

[2] Gifu Shotoku Gakuen University, Nakauzura 1-38, Gifu-shi, Gifu, 501-6194 Japan

ttakahashi@gifu.shotoku.ac.jp

Abstract

In this paper, we propose a method for recognizing low-resolution characters using a super-resolution technique. Although portable digital cameras can be used for camera-based character recognition, the captured images contain several types of noises which make the recognition task difficult. We introduce a phase of super-resolution before the recognition to enhance the resolution of images obtained from a video. The proposed method uses the subspace method for the recognition of characters which are integrated from multiple low-resolution characters by the super-resolution technique. Experimental results show that the proposed method improves the recognition accuracy; we confirmed that the recognition rate for the input size of 7×7 pixels was 90.35%, and for the input size of 9×9 pixels was 99.97%.

1. Introduction

Portable digital cameras are widely used in recent years. Character recognition techniques using these cameras have come into practical use. It enables us to recognize characters simply by capturing the document. For example, we can scan and input personal information from business cards or URLs from magazines. Although many character recognition methods have already been proposed, it is still difficult to recognize them accurately when a target document contains many characters. This is because the resolution of each character becomes lower in the case that many characters are included in the target text region (Fig. 1). Elms et al. [1] have proposed a method to recognize low-resolution characters captured by an image scanner. However, the performance of this method is insufficient to recognize characters when the resolution of a single character is low. Against this problem, methods for integrating multiple low-resolution images obtained from a video have been proposed. Yanadume et al. [2] have pro-
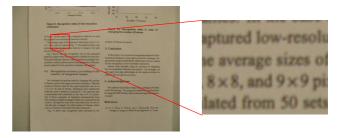

Figure 1. Example of low-resolution characters obtained from a relatively large target text region.

posed to integrate a sequence of similarity values given by a subspace method [3, 4, 5, 6] for the classification. It is based on the idea that the use of multiple images can restrict mis-classification. However, when multiple images include many images closer to an incorrect category, it is difficult to restrict misclassifications.

As an alternative approach to low-resolution character recognition in a video, we propose a method that uses a super-resolution technique. The recognition scheme of the proposed method is presented in Fig. 2. The proposed method reconstructs a high-resolution image from multiple low-resolution images by means of the super-resolution technique, in order to enhance the resolution of both training data and test data images. Hence, we expect to recognize the test data correctly.

This paper is organized as follows: Section 2 introduces the super-resolution technique used to enhance the resolution of the character images. In Section 3, the recognition step using the subspace method is described. Experimental results are presented in Section 4.

2. Super-resolution technique

An image captured by a camera is generally influenced by several noises (for example, hand movement, optical blur and so on). These influences are usually modeled by a PSF
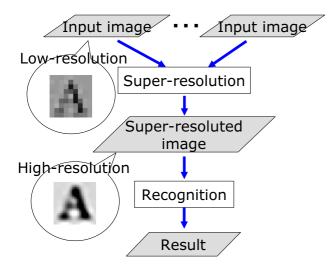
Figure 2. The recognition scheme of the proposed method.

(a) Low-resolution images of 'A'

(b) Low-resolution images of 'B'

(c) Image super-resoluted from (a)

(d) Image super-resoluted from (b)

Figure 3. Examples of character 'A' and 'B' reconstructed by super-resolution.
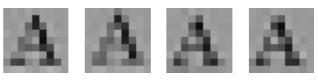
(Point Spread Function). Moreover, when a full document is captured in a single shot, the image of each character tends to be too small for recognition. The super-resolution technique reconstructs a high-resolution image from multiple low-resolution images which have different subpixel shifts from each other [7, 8, 9, 10, 11].

Given low-resolution character images obtained from a sequence of video frames captured by a video camera, the method first integrates the low-resolution images into a high-resolution image by subpixel matching. Subpixel shifts from each image are calculated by matching the images resampled from integer pixels to subpixels. Then, according to subpixel shifts, the resampled low-resolution images are integrated into a high-resolution image.
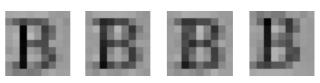
Next, a target image is updated iteratively until the evaluation function I defined in Eq. 1 becomes sufficiently small, which removes the influences of optical blur in the target image.

$$I = \sum_{y=1}^{H} \sum_{x=1}^{W} \left[\boldsymbol{b}(x,y)^T \boldsymbol{h} - f(x,y) \right]^2. \qquad (1)$$

Here $\boldsymbol{h}$ is the target image in the vector form, $\boldsymbol{b}(x, y)$ is a vector of the PSF kernel at a position (x, y), and $f(x, y)$ is the pixel value of the integrated image at the position. H and W are the height and the width of the image, respectively. The evaluation function I is minimized by the conjugate gradient method proposed by Fletcher and Reevess [12]. Fig. 3 shows examples of high-resolution character images which are reconstructed by the super-resolution technique.

3. Recognition by the subspace method

We employ the subspace method, which is robust against slight transformation of patterns, for the recognition step of the proposed method. The subspace method consists of two stages; a learning stage and a recognition stage. The learning stage constructs a subspace for each category, and then the recognition stage calculates similarities between an input image and the subspaces.

3.1. Training data

Training data are reconstructed from multiple-frame images by the super-resolution technique. We use a sequence of multiple frame images from a video which captures a printed list of alphabets as training data to obtain character images with plenty of variations for each low-resolution character.

3.2. Construction of a subspace

A subspace approximates a distribution of the training data. The learning stage finds orthogonal bases that well approximate the distribution of the training data for each category. Each i-th training image is normalized to a unit vector whose mean value is equal to 0. The normalized

(a) The first eigenvector   (b) The second eigenvector   (c) The third eigenvector

Figure 4. Eigenvectors obtained from 100 samples of character 'A'.

vector is represented as

$$\boldsymbol{x}_i = [x_1, x_2, \cdots, x_N]^T, \tag{2}$$

where N is the number of pixels. Next, a matrix $\mathbf{X}$ is defined as

$$\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_K], \tag{3}$$

where K is the number of training data used for a category. Then an autocorrelation matrix $\mathbf{Q}$ is calculated from the matrix $\mathbf{X}$ by

$$\mathbf{Q} = \mathbf{X}\mathbf{X}^T. \tag{4}$$

A subspace is constructed for each category as R ($\leqq K$) eigenvectors that correspond to the largest R eigenvalues. Each r-th eigenvector is represented as $\boldsymbol{e}_r^{(c)}$, where c ($= 1, 2, \cdots, C$) represents a category. Fig. 4 shows examples of eigenvectors.

3.3. Recognition

Each input character image is segmented from high-resolution images which are reconstructed by the super-resolution technique. Let $\boldsymbol{y}$ be an input image which are normalized in the same manner as in the learning stage. The similarity between a category c and the input image $\boldsymbol{y}$ is defined as

$$L^{(c)}(\boldsymbol{y}) = \sum_{r=1}^{R} \left(\boldsymbol{e}_r^{(c)} \cdot \boldsymbol{y} \right)^2, \tag{5}$$

where $(\boldsymbol{a} \cdot \boldsymbol{b})$ represents the inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$. Then the input image is classified to a category c which maximizes Eq. (5). The category c is calculated by the following equation

$$\hat{c} = \arg\max_c L^{(c)}(\boldsymbol{y}). \tag{6}$$

4. Experiments and disscussion

We confirmed the effectiveness of our method experimentally. Test samples were obtained by capturing a video sequence of printed characters with a portable digital camera (Table 1). Fig. 5 shows an example of the images that

Table 1. Specification of the digital camera.

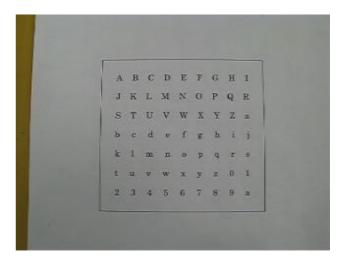| Product model	Canon PowerShot G9
Resolution	360×240 pixels
Frame rate	30 fps

Figure 5. Example of a captured image.

were used in the experiment. During this experiment, because subpixel shifts between the captured frames is essential to the proposed method, the anti-blur function of the digital camera was kept off. An alphanumeric "Century" font was used in the experiments. The number of categories C was 62 as shown in Fig. 6. In applying the super-resolution technique, we assumed that the PSF kernel could be represented by a Gaussian filter ($\sigma = 0.3$). When we constructed a subspace, the number of eigenvectors R was set to 5, and the size $W \times H$ of images was normalized to 32×32 pixels.

4.1. Recognition accuracy by image sizes

We compared recognition rates of the proposed method to a comparative method by changing the size of characters. In the comparative method, characters are recognized from multiple low-resolution images without applying the super-resolution technique. The comparative method obtains a recognition result $\hat{c}$ by the following equation

$$\hat{c} = \arg\max_c \sum_{m=1}^{M} L'^{(c)}(\boldsymbol{y}_m) \tag{7}$$

where $\boldsymbol{y}_m$ represents the m-th input of a low-resolution image, M is the number of input images, $L'^{(c)}(\cdot)$ is the similarity to category c's subspace constructed from c's low-resolution images. The proposed method employed the super-resolution technique for the recognition.
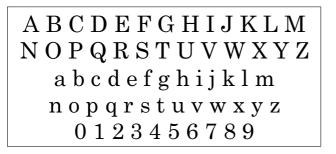
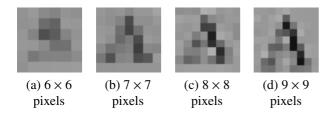**Figure 6.** Character categories (Century font).

| (a) 6×6 pixels | (b) 7×7 pixels | (c) 8×8 pixels | (d) 9×9 pixels |

Figure 7. Example of test data for each size.

Figure 8. Recognition rates of low-resolution characters. In the proposed method, a character was super-resoluted from 30 frames. In the comparative method, 30 low-resolution characters were used directly.

| (a) 2 frames | (b) 8 frames | (c) 32 frames |

Figure 9. The super-resoluted characters.

As training data, 100 characters per category (a total of 6,200 images) were prepared for each method. In the case of the proposed method, characters were super-resoluted from 30 frames. In the case of the comparative method, the captured low-resolution characters were used directly.

The average sizes of the captured characters were 6×6, 7×7, 8×8, and 9×9 pixels (Fig. 7). Recognition rates were calculated from 50 sets of test data per category for each size and each method.

Fig. 8 shows that the recognition rate of the proposed method is higher than the comparative method for any size. This result indicates that the super-resolution technique successfully worked, and also proved the superiority of the proposed method for low-resolution character recognition.

4.2. Recognition accuracy by the number of images used for super-resolution

We also evaluated the recognition rates by changing the number of frames used for the super-resolution; 1, 2, 4, 8, 16, and 32 frames. Subspaces were constructed under the same condition as in Section 4.1. We used test data reconstructed from characters with the size of 8×8 pixels. Fig. 9 shows examples of characters reconstructed from low-resolution images. Recognition rates were calculated from 50 sets of test data per category for each number of frames, where each set consists of reconstructed input characters.

Fig. 10 shows that the recognition rate increases in proportion to the number of frames used for the super-resolution. This result indicates that the super-resolution technique complements insufficient resolution to recognize characters by integrating multiple low-resolution images.
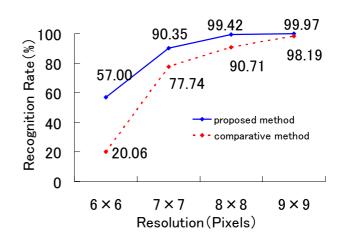
4.3. Disscussion

From Fig. 8 and Fig. 10, we confirmed that the proposed method could recognize characters accurately, compared to the conventional method. However, the recognition accuracy declines in proportion to the input character size. As shown in Fig. 11, when a character was super-resoluted from an image size of 9×9 pixels, even similar characters 'I' and '1' could be distinguished. Meanwhile, when super-resoluted from 7×7 pixels, it is difficult to distinguish between them. We consider that such phenomena caused the decline of the recognition rate in low-resolutions in Fig. 8.

5. Conclusion

In this paper, we proposed a recognition method for low-resolution characters using the super-resolution technique. Experimental results exhibited the effectiveness of our method for the recognition of low-resolution characters.

Future works include using the structure of characters for low-resolution character recognition. For example, we will
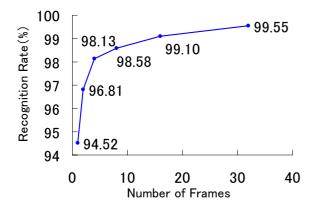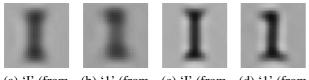
Figure 10. Recognition rates by changing the number of frames used for the super-resolution.

(a) 'I' (from images of 7×7 pixels) (b) '1' (from images of 7×7 pixels) (c) 'I' (from images of 9×9 pixels) (d) '1' (from images of 9×9 pixels)

Figure 11. Example of super-resoluted images of similar characters.

apply edge information to the super-resolution to improve the recognition accuracy.

6. Acknowledgement

References

[1] A. J. Elms, S. Procter, and J. Illingworth, "The advantage of using an HMM-based approach for faxed word recognition," Int. J. Document Analysis and Recognition, vol.1, no.1, pp.18–36, February 1998.

[2] S. Yanadume, Y. Mekada, I. Ide, and H. Murase, "Recognition of very low-resolution characters from motion images captured by a portable digital camera," Proc. 2004 Pacific-Rim Conf. on Multimedia, Lecture Notes in Computer Science, vol.3332, pp.489–496, Springer-Verlag, December 2004.

[3] E. Oja, "Subspace methods of pattern recognition," Research Studies, Hertfordshire, UK, December 1983.

[4] H. Murase, H. Kimura, M. Yoshimura, and Y. Miyake, "An improvement of the auto-correlation matrix in the pattern matching method and its application to hand-printed "HIRAGANA" recognition (in Japanese)," IECE Trans., vol.J64-D, no.3, pp.276–283, March 1981.

[5] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," Int. J. Computer Vision, vol.14, no.1, pp.5–24, January 1995.

[6] S. Omachi and H. Aso, "A qualitative adaptation of subspace method for character recognition," Systems and Computers in Japan, vol.31, issue.12, pp.1–10, Wiley, November 2000.

[7] C. P. Sung, K. P. Min, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," IEEE Signal Process. Mag., vol.20, no.3, pp.21–36, May 2003.

[8] M. Tanaka and M. Okutomi, "A fast algorithm for reconstruction-based super-resolution and its accuracy evaluation," Systems and Computers in Japan, vol.38, issue.7, pp.44–52, Wiley, June 2007.

[9] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," Proc. 1995 IEEE Int. Conf. on Image Processing, vol.2, pp.539–542, October 1995.

[10] R. R. Schulz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. Image Process., vol.5, no.6, pp.996–1011, June 1996.

[11] H. Stark and P. Oskoui, "High resolution image recovery from image-plane arrays, using convex projections," J. Opt. Soc. Am. A, vol.6, pp.1715–1726, November 1989.

[12] R. Fletcher and C. Reeves, "Function minimization by conjugate gradients," Computer Journal, vol.7, no.2, pp.149–154, British Computer Society, 1964.